Identifying Active Factors from Non-Geometric Plackett-Burman Designs and their Half-Fractions

Oddgeir Samset
Department of Mathematical Sciences
The Norwegian University of Science and Technology
N-7034 Trondheim, Norway

January 26, 1999

Abstract

The non-geometric Plackett-Burman designs open for an economical screening of many factors in few runs. These designs have traditionally been used as main effect plans assuming interactions to be negligible. If interactions are present, however, a more thorough analysis of the experiments is required to identify the active factors due to the complex alias-patterns of these designs. In this paper we demonstrate the usefulness of two variable selection techniques that account for interactions. One is a best subsets regression procedure guided by the principle of effect heredity. The other is a stepwise regression procedure on subsets of potentially active factors. Both techniques enable a fairly systematic search for the active factors within reasonable computing time.

Key words: Effect heredity; Factor sparsity; Regression analysis; Screening; Variable selection.

1 Introduction

Screening experiments play an important part of experimental design. Their primary objective is to identify the *active* factors, i.e. factors that individually or interactively give rise to the main variation in the response. Hence focus can be placed on these factors to obtain a more efficient and economical future experimentation. Factors that not are active are said to be *inert*. These are often important from another point of view. They indicate flexibility of the process or system under study since their levels to some extent can be adjusted freely to their most appropriate values without consequences for the given response.

In screening situations one typically experiences that only a few of the factors considered initially are active while most of the factors are inert. This is called the *principle of factor sparsity* by Box and Meyer (1986), and is one of the rationales for using highly fractionated designs such as non-replicated 2^{k-p} fractional factorial designs (Finney, 1945; Box and Hunter, 1961) and Plackett-Burman designs (Plackett and Burman, 1946) for screening. These designs have nice properties when projected onto a few factors. Thus, under factor sparsity they are regarded well suited for a systematic search for the active factors.

It has recently been an increased interest for the non-geometric Plackett-Burman (PB) designs and how to analyse experiments from these designs. There are three main reasons for this. First, PB designs give the opportunity to study $n-1$ factors in n runs where n is a multiple of four. When n is a power of two, the PB designs coincide with the 2^{k-p} fractional factorial (FF) designs. The other PB designs, often referred to as *non-geometric*, therefore open for an economical screening of many factors in few runs. Secondly, computer search (Lin and Draper, 1992; Box and Bisgaard, 1993) showed that saturated non-geometric PB designs might have better projection properties than saturated FF designs. See also Box and Tyssedal (1996) and Tyssedal and Samset (1997). Thirdly, as pointed out by Hamada and Wu (1991), non-geometric PB designs and other designs with partially confounded effects give under factor sparsity the ability of simultaneously estimating both main effects and interactions. Thus, unlike the FF designs where fully confounded effects may result in many models, non-geometric PB designs have in certain situations potential for separating out one plausible model that explains the main variation in the response.

The partial confounding of effects can be viewed as a potential of the non-geometric PB designs, but it is also obvious that this makes the analysis of the experiments more troublesome. Section 2 first gives a short overview of existing methods for analysing experiments from such designs. Then two additional variable selection techniques are outlined. One is a best subsets regression procedure guided by the principle of effect heredity. The other is a stepwise regression procedure on subsets of potentially active factors. Both methods enable a fairly systematic search for the active factors within reasonable computing time. Only main effects and two-factor interactions are here taken into account. The methods described can however easily be extended to also include interactions of higher order or quadratic terms. In Section 3 the two methods are demonstrated for screening 11 and 22 factors using the 12 run PB design and the half-fraction of the 24 run PB design, respectively. The response values are simulated from a model with four active factors introduced by Hamada and Wu (1992).

Focus is here placed on analysing experiments from non-geometric PB designs and their half-fractions. However, these methods can also be applied to other designs with complex alias-patterns. This can be multi- and mixed-level designs, and situations where there are hard-to-control factors.

2 Identifying the Active Factors

Experiments from FF designs have traditionally been analysed using normal or half-normal plot (Daniel, 1959; 1976) or Bayes plot (Box and Meyer, 1986). These methods are also often used for analysing non-geometric PB designs assuming interactions between factors to be negligible. However, presence of active interactions can seriously distort the traditional interpretation of such plots due to the complex alias-patterns of these designs. Hence, active factors may wrongly appear as inert and vice versa.

A good illustration of this problem is *example 4* of Hamada and Wu (1992). The response values are simulated from the model $y = x_1 + 2x_1x_2 + 2x_1x_3 + \epsilon$ with $\epsilon \sim N(0, 0.25^2)$ using the 12 run PB design in Table 1. A realisation of the response values from this model is given by

$$y = (1.08\ 0.87\ \text{-}5.00\ 1.40\ 1.21\ 5.07\ \text{-}5.23\ \text{-}1.02\ 3.04\ \text{-}3.14\ \text{-}1.08\ 3.10\ )'.$$

Figure 1 shows the half-normal plot of the estimated main effect contrasts for these response values. Using a traditional interpretation of the half-normal plot, the four largest estimated contrasts seem to be due to the inert factors x_7, x_8, x_9 and x_{10}. However, it is not difficult to verify that these large estimated contrasts are a result of the complex alias-pattern of the design, see Tyssedal and Samset (1997). Thus, in this and similar situations the active factors can easily be overlooked.

A number of methods have recently been proposed for analysing non-geometric PB and other designs with complex alias-patterns when interactions are taken into account. Hamada and Wu (1992) proposed a method based on iterative use of forward selection. Box and Meyer (1993) presented a Bayesian approach that based on the principle of factor sparsity estimates the marginal posterior probabilities that the individual factors are active. Tyssedal and Samset (1997) used an alternative interpretation of the half-normal plot where the alias-pattern of the applied design is taken into account. Chipman, Hamada and Wu (1997) applied a Bayesian approach using the SSVS (stochastic search variable selection) algorithm of George and McCulloch (1993) and Gibbs sampling.

2.1 Heredity Least Term Regression

Best subsets regression can in principle also be used for analysing experiments from non-geometric PB designs. The advantage with best subsets regression is that it examines all possible combinations of terms. Hence, this method will inevitable identify the 'best' model if it is distinguishable from the rest according to some criteria such as the multiple coefficient of determination R^2, R^2_{adj}, Mallow's C_p or the residual mean square error s^2. If there are more models that equally well explain the variation in the response, all of these models will be identified. However, a relatively large number of factors are typically of interest in screening situations. When both main effects and interactions are taken into account, the number of models to examine may therefore grow very large. In Table 2 we have listed the number of models to examine for a given number of terms (main effects and/or two-factor interactions in addition to the intercept) for the saturated 12, 24 and 48 run PB designs. Another problem with best subsets regression is that a large proportion of the models examined only contain interactions. These models are often regarded less plausible for explaining the data.

The *principle of effect heredity* (Hamada and Wu, 1992) has been used for a long time in the analysis of experiments from FF designs to indicate the most plausible models, see for example the *Injection molding example* in Box, Hunter and Hunter (1978, p. 399). This principle states that in order to include an interaction, at least one of the corresponding main effects should also be in the model. Thus, an alternative approach to best subsets regression is to guide the inclusion of terms into the models according to the principle of effect heredity. The methodology is hence similar to best subsets regression, but the models are restricted to the following types:

No term: $E(y) = \beta_0$

One term: $E(y) = \beta_0 + \beta_1 x_i \quad i = 1, \dots, p$

Two terms: $E(y) = \beta_0 + \beta_1 x_i + \beta_2 x_j \quad i = 1, \dots, p-1; \; j > i$

$E(y) = \beta_0 + \beta_1 x_i + \beta_{12} x_i x_j \quad i = 1, \dots, p-1; \; j \neq i$

Three terms: $E(y) = \beta_0 + \beta_1 x_i + \beta_2 x_j + \beta_3 x_k \quad i = 1, \dots, p-2; \; j > i; \; k > j$

$E(y) = \beta_0 + \beta_1 x_i + \beta_2 x_j + \beta_{13} x_i x_k \quad i = 1, \dots, p-1; \; j > i; \; k \neq i$

$E(y) = \beta_0 + \beta_1 x_i + \beta_2 x_j + \beta_{23} x_j x_k \quad i = 1, \dots, p-1; \; j > i; \; k \neq (i,j)$

$E(y) = \beta_0 + \beta_1 x_i + \beta_{12} x_i x_j + \beta_{13} x_i x_k \quad i = 1, \dots, p-1; \; j \neq i; \; k \neq (i,j)$

In Table 3 we have listed the number of models to examine for a given number of terms (main effects and/or two-factor interactions in addition to the intercept) for the saturated 12, 24 and 48 run PB designs. Compared to best subsets regression the number of models to examine is now significantly reduced.

As pointed out by Meyer, Steinberg and Box (1996), effect heredity is a principle that not always can be relied on, especially if qualitative factors are present. Effect heredity should hence not be forced into the models. It is therefore important to notice that the principle of effect heredity only guides the inclusion of terms into the models for the method described here. Models that do not follow this principle can still be found when including additional main effect terms into the models. That is, if the response is given by $E(y) = \beta_0 + \beta_{12}x_1x_2$ this model will be among the two-term model as $E(y) = \beta_0 + \beta_1 x_1 + \beta_{12}x_1x_2$ or $E(y) = \beta_0 + \beta_2 x_2 + \beta_{12}x_1x_2$ with $\hat{\beta}_1$ and $\hat{\beta}_2$ close to zero. Such models need extra terms and more models have to be examined compared to if effect heredity holds. Nevertheless, it seems reasonable to give higher priority to the most rational models and examine these models first.

2.2 Stepwise Submodel Regression

We will here consider a method that to some extent combines the 'factor based' strategy proposed by Box and Meyer (1993) and the 'term based' strategy using forward regression proposed by Hamada and Wu (1992). A detailed description of the method in a more general context is given by Samset (1998).

When there are a large number of factors of interest, and when both main effects and interactions are taken into account, then the total number of terms is large. A problem with stepwise regression is that as the number of terms grow large the probability of choosing inert terms typically increases. It is therefore reasonable to somehow restrict the number of terms for which we perform stepwise regression. Hamada and Wu (1992) address this problem by first using a traditional analysis such as half-normal plot to identify the large main effect contrasts. Forward regression is then used for choosing among these main effect contrasts and all two-factor interaction contrasts containing at least one of the corresponding main effect factors. To our experience, however, there are two problems with this approach. First, if two-factor interactions are active, inert main effect contrasts may wrongly appear as active and vice versa due to the complex alias-pattern when using a traditional analysis method such as half-normal plot. Thus, there is a danger that focus initially is placed on inert terms. Their example 4, see Figure 1, is a good illustration of this problem. Secondly, considering all two-factor interactions containing at least one factor corresponding to the large main effect contrasts may still result in a relatively large number of terms. With three large main effect contrasts found initially, as many as 27 additional two-factor interaction contrasts are to be considered by the forward procedure for the saturated 12 run PB design.

One alternative approach to restrict the number of terms is to divide the factor space according to the principle of factor sparsity. The motivation is that by assuming none, one, two and so on active factors, one has the opportunity to perform a systematic search for every subset of potentially active factors. This 'factor based' strategy was proposed by Box and Meyer (1993). Box and Meyer include all terms related to a subset of potentially active factors into the model, and use a Bayesian approach for computing the posterior probabilities of the model given the observed response. Due to the relatively few number of runs typically available for a screening design, it seems reasonable to remove extraneous terms from the models, see also Samset and Tyssedal (1998b).

Samset (1998) proposed to use stepwise regression on the subsets of terms based on factor sparsity. These subsets are of the following types:

One active factor: $\{x_i\}$ $i = 1, \ldots, p$
Two active factors: $\{x_i, x_j, x_i x_j\}$ $i = 1, \ldots, p-1;\ j > i$
Three active factors: $\{x_i, x_j, x_k, x_i x_j, x_i x_k, x_j x_k\}$ $i = 1, \ldots, p-2;\ j > i;\ k > j$

This methodology gives a fairly systematic search for the active factors. One will always examine a subset that contains the active factors, and hence all the active main effect and two-factor interaction terms. If the assumption of factor sparsity holds, the number of potentially active factors to include, and the number of models to examine, will be low. Then there also are relatively few terms within each subset. Stepwise regression takes the 'level of significance' of the individual terms in the *submodels* into account, and will therefore help preventing extraneous terms. This again increases the precision and prevents over-fitting.

Samset (1998) use the difference in the multiple coefficient of determination,

$$\Delta R^2_{\mathrm{IN/OUT}} = R_i^2 - R_{i-1}^2 = SS(\beta_i | \beta_{i-1}, \ldots, \beta_0)/SS_T$$

as a stopping criterion instead of the traditionally $F_{\mathrm{IN/OUT}}$. The motivation is that $\Delta R^2_{\mathrm{IN/OUT}}$ is proportional to $SS(\beta_i|\beta_{i-1}, \ldots, \beta_0)/\sigma^2$. Since σ^2 is unknown we can not find the proportionality constant $K = \sigma^2/SS_T$. Note however that K is the same for all submodels we are examining. The procedure is then to perform stepwise regression for each of the subsets of terms defined above, and vary the value of $\Delta R^2_{\mathrm{ENTER/REMOVE}}$ until one or a few models separate out from the rest. Hence this criterion enables all submodels to be compared at the same 'level of significance' as if σ^2 was known. For the subset containing the active factors stepwise regression hopefully will perform well under factor sparsity due to the low correlation between the different terms typically present in a designed experiment. For the other subsets, however, the restrictions on the terms to chose from will make it more difficult to find models that fit the data well. Hence, the restrictions on the terms also here tend to put higher priority on the plausible models.

The focus here is on stepwise regression. However, due to the rather low correlation between the contrast columns for a designed experiment, forward regression is often appropriate. One advantage by using forward compared to stepwise regression is of course the possibility of reducing the computing time. For forward regression there is no need for recomputing any of the results even if the value of $\Delta R^2_{\mathrm{ENTER}}$ is varied. In Table 4 we have listed the maximum number of models that may have to be computed using forward regression, i.e. the number of submodels multiplied by the number of terms in a submodel, for the saturated 12, 24 and 48 run PB designs. Compared to best subsets and heredity least term regression the number of models to examine is substantially reduced. In most situations the forward procedure terminates long before all terms in each submodel are included. Thus the number of models to examine is usually much smaller than the values listed.

3 Examples

3.1 Example 1

We will here consider *example 5* from Hamada and Wu (1992). The response values are simulated from the model $y = 2x_1 + 4x_3 + 2x_2x_3 - 2x_3x_4 + \epsilon$ with $\epsilon \sim N(0, 0.5^2)$ using the 12 run PB design in Table 1. We use a realisation of the response values from this model given by Hynén (1996)

$$y = (6.27\ \text{-}3.21\ 5.51\ 1.94\ \text{-}2.48\ 10.35\ 2.47\ \text{-}2.84\ \text{-}2.93\ \text{-}1.19\ \text{-}10.09\ \text{-}5.98)'.$$

Both Hamada and Wu (1992) using their proposed forward selection technique, Hynén (1996) using the Bayesian method of Box and Meyer (1993), and Chipman et al. (1997) using the stochastic search variable selection algorithm, have problems identifying the active factors in this example, and indicate that additional experiments are required.

In Figure 2 to Figure 4 the three models with smallest residual mean square error are listed for zero to four terms for best subsets regression (BSR) and heredity least terms regression (HLTR), and for zero to four potentially active factors for forward submodel regression (FSR). From the results for BSR we see that the correct model, and hence the active factors, is clearly identified when including four terms. Thus, additional runs are not required to identify the correct model in this situation when we are using the 12 run PB design. Also HLTR and FSR with $\Delta R^2_{\text{ENTER}} = 6.0$ identify the correct model using considerably less computer time than best subsets regression. Observe that while BSR examines as many as 768692 models, HLTR and FSR examine only 17392 and 2262 models, respectively.

When using highly fractionated designs such as non-geometric PB designs in screening situations, there is of course no guarantee that the design enables identification of a single model that separates out from the rest according to some criteria. One problem is a large noise compared to the signal in the response. To study the effect of an increasing value of the standard deviation of the noise, σ, additional 100 realisations of the response values from the model described above are generated for different values of σ. In Table 5 we have listed the percent cases where the correct model is rated first, second, third and so on among all four-term models for HLTR and four active factors for FSR for σ equal to 0.5, 1.0 and 1.5. A problem with FSR is that the stopping criterion $\Delta R^2_{\text{ENTER}}$ needs to be tuned to obtain a good result. This tuning process is here simplified by first performing forward regression on the subset containing the four active factors, that is the ten terms $x_1, \ldots, x_4, x_1x_2, \ldots, x_3x_4$. Then the minimum value of ΔR^2_{IN} for the first four terms added into this submodel is used as $\Delta R^2_{\text{ENTER}}$. This is a 'optimal' tuning in the sense that if we use a larger value of $\Delta R^2_{\text{ENTER}}$ only three or less terms enter the model, and hence the correct model are not found. In addition to the tuning we have also used a constant value $\Delta R^2_{\text{ENTER}} = 5.0$ to give an impression of how critical the optimal tuning is.

For $\sigma = 0.5$ the correct model is rated first among all the four-term models in 100% of the realisations using HLTR. For tuned FSR the correct model is rated first for 46%, and not less than third for 97% of the realisations. The sample mean and standard deviation of the tuned $\Delta R^2_{\text{ENTER}}$ are 6.8 and 1.1, respectively. Even with $\Delta R^2_{\text{ENTER}} = 5.0$ the correct model is rated among the three best for 84% of the situations. For two cases, indicated within parentheses, the correct model is not found among the first 10 models since the stopping criterion then was specified too large. Also for $\sigma = 1.0$ the proposed methods perform very well. For HLTR the correct model is rated first for 93%, and not less than third for 99% of the realisations.

For tuned FSR the correct model is not rated less than third for 93%. The sample mean and standard deviation of the tuned $\Delta R^2_{\text{ENTER}}$ are here 7.1 and 1.9, respectively. Actually, for both $\sigma = 0.5$ and $\sigma = 1.0$ the situations for FSR are better than the impression we get from the table since the other models with higher rating than the correct model typically tend to have a relatively large number of terms. Hence the correct model stand out from the rest even though it is not rated first. For $\sigma = 1.5$ HLTR now start having problem identifying the correct model. Even though the correct model is rated among the six first for as many as 87% of the realisations, the 'best' models typically have rather similar values of s. Thus there seem to be a need for additional experiments in order to further separate between the models when using the proposed methods. Having identified the competing models, the gain in information from the experiment might nevertheless be large. This information can be used for improved planning of follow-up experiments to resolve the ambiguities, see Meyer et al. (1996).

3.2 Example 2

Lin (1993) proposed to use half-fractions of the non-geometric PB designs as super-saturated designs for screening more than $n-1$ factors in n runs. Inspired by the performance of the two methods in the previous example they are also applied for identifying the active factors from such designs. To our knowledge these designs have only been regarded as main effect plans when interactions are assumed negligible (Lin, 1993; Chipman et al., 1997; Abraham, Chipman and Vijayan, 1998; Westfall, Young and Lin, 1998).

The same model as above is used, i.e. the response values are simulated from the model $y = 2x_1 + 4x_3 + 2x_2x_3 - 2x_3x_4 + \epsilon$ with $\epsilon \sim N(0, 0.5^2)$, but now using a half-fraction of the 24 run PB design to screen 22 factors in 12 run. The design is given in Table 6. The realisations of the noise are here given by

$$\epsilon = (\text{-0.13 0.90 -0.32 0.05 -0.36 0.21 -0.97 0.33 -0.55 -0.05 -0.53 -0.62})'.$$

The results from heredity least term regression (HLTR) and forward submodel regression (FSR) are given in Figure 5 and Figure 6, respectively. For both methods the correct model clearly separates out as a plausible candidate. Observe that among the 334610 models examined by HLTR there are 132 situation where the matrix of regression variables (including the intercept) is not of full rank. In these situation the Moore-Penrose pseudo-inverse are used to find a solution to the normal equations.

Also here additional 100 different realisations of the response values are generated for σ equal to 0.5, 1.0 and 1.5. The results are given in Table 7. For $\sigma = 0.5$ the correct model is also here rated first among all four-term models in 100% of the realisations using HLTR. For tuned FSR the correct model is now rated first for 85%, and not less than third for 100% of the realisations. Somewhat surprisingly it seems that FSR perform better here that in the previous example. However, the sample mean and standard deviation of the tuned $\Delta R^2_{\text{ENTER}}$ are 8.3 and 1.0, respectively. A higher value of $\Delta R^2_{\text{ENTER}}$ will remove more of the other models, and the correct model thus get a higher ranking. For $\sigma = 1.0$ the correct model is rated first in 56% and not less than third in 79% of the realisations using HLTR. For tuned FSR the correct model is now rated first for 37%, and not less than third for 52% of the realisations.

Draper and Lin (1995) used the *repeat and mirror-image pattern* to characterise the different projections of a design. For the design in Table 6 there are 22 different repeat and mirror-image patterns for projections into 4 dimension, see Samset and Tyssedal (1998a). For the columns (1, 2, 3 and 4) this pattern is represented as 1[2 1] 3[1 1] 3[1 0]. One run have both a repeated (there are two repeats) and a mirror-image run, three runs have no repeated but a mirror-image run, and 3 runs have neither repeated nor mirror-image runs. Thus, 11 of the possible 16 unique runs in the full 2^4 design are present. This is one of the better projections into 4 dimension for the half-faction of the 24 run PB design. For the columns (1, 2, 5 and 6) the corresponding repeat and mirror-image pattern is given by 2[3 0] 6[1 0]. Two runs have two repeated and no mirror-image runs, and 6 runs have neither repeated nor mirror-image runs. This shows that only 8 of the 16 unique runs are present. Since there are less unique runs and no mirror-image runs this can be considered a more complicated situation. Thus, if the active factor are x_1, x_2, x_5 and x_6 the correct model should now probably be worse to separate out from the rest. We have therefore also considered the model $y = 2x_1 + 4x_5 + 2x_2x_5 - 2x_5x_6 + \epsilon$ with $\epsilon \sim N(0, 0.5^2)$. The same realisations of the noise as before are used. There are now three models that separate out from the rest for HLTR, and all these models contain the four active factors. It is easy to verify that the model $E(y) = 2x_5 - 2x_1x_2 - 2x_5x_6$ also gives the same signal as the correct model. This three-term model can not be found among models with three terms since it does not follow the principle of effect heredity. It is rather found among the four-term models with either x_1 or x_2 added. By using FSR the three-term models is identified.

In Table 8 additional 100 different realisations of the response values are generated for σ equal to 0.5, 1.0 and 1.5. Now the minimum value of ΔR^2_{IN} for the first three terms added into the submodel with the active factors is used as $\Delta R^2_{\mathrm{ENTER}}$. Again both HLTR and FSR perform very well when $\sigma = 0.5$, and the correct model is rated among the three first in 100% of the realisations. However, the effect of increasing σ is now larger. Even for $\sigma = 1.00$ the correct model is not found among the 10 first for 23% and 22% of the realisations, respectively.

References

Abraham, B., Chipman, H. and Vijayan, K. (1998). Another Look at Supersaturated Designs, *Technical Report RR-98-03*, The Institute for Improvement in Quality and Productivity. University of Waterloo, Canada.

Box, G. E. P. and Bisgaard, S. (1993). What Can You Find Out from 12 Experimental Runs, *Quality Engineering* **5**, 663–668.

Box, G. E. P. and Hunter, J. S. (1961). The 2^{k-p} Fractional Factorial Designs, *Technometrics* **3**, 311–351, 449–458.

Box, G. E. P., Hunter, W. G. and Hunter, J. S. (1978). *Statistics for Experimenters*, John Wiley & Sons, New York.

Box, G. E. P. and Meyer, R. D. (1986). An Analysis for Unreplicated Fractional Factorials, *Technometrics* **28**, 11–18.

Box, G. E. P. and Meyer, R. D. (1993). Finding the Active Factors in Fractionated Screening Experiments, *Journal of Quality Technology* **25**, 94–105.

Box, G. E. P. and Tyssedal, J. (1996). Projective Properties of Certain Orthogonal Arrays, *Biometrica* **83**, 950–955.

Chipman, H., Hamada, M. and Wu, C. F. J. (1997). A Bayesian Variable-Selection Approach for Analyzing Designed Experiments With Complex Aliasing, *Technometrics* **39**, 372–381.

Daniel, C. (1959). Use of Half-Normal Plots in Interpreting Factorial Two-Level Experiments, *Technometrics* **1**, 311–341.

Daniel, C. (1976). *Applications of Statistics to Industrial Experimentation*, John Wiley & Sons, New York.

Draper, N. R. and Lin, D. K. J. (1995). Characterizing Projected Designs: Repeat and Mirror-Image Runs, *Commun. Statist. Theory Meth.* **24**, 775–795.

Finney, D. J. (1945). The Fractional Replication of Factorial Arrangements, *Annals of Eugenics* **12**, 291–301.

George, E. I. and McCulloch, R. E. (1993). Variable Selection Via Gibbs Sampling, *Journal of the American Statistical Association* **88**, 881–889.

Hamada, M. and Wu, C. F. J. (1991). Analysis of Censored Data From Highly Fractionated Experiments, *Technometrics* **33**, 25–38.

Hamada, M. and Wu, C. F. J. (1992). Analysis of Designed Experiments with Complex Aliasing, *Journal of Quality Technology* **24**, 130–137.

Hynén, A. (1996). Screening for Main and Interaction Effects with Plackett and Burman's 12 Run Design, *Technical Report 1*, Division of Quality Technology and Management. Lindköping University, Sweden.

Lin, D. K. J. (1993). A New Class of Supersaturated Design, *Technometrics* **35**, 28–31.

Lin, D. K. J. and Draper, N. R. (1992). Projection Properties of Plackett and Burman Designs, *Technometrics* **34**, 423–428.

Meyer, R. D., Steinberg, D. M. and Box, G. E. P. (1996). Follow-Up Designs to Resolve Confounding in Multifactor Experiments (with discussion), *Technometrics* **38**, 303–332.

Plackett, R. L. and Burman, J. P. (1946). The Design of Optimum Multifactorial Experiments, *Biometrica* **33**, 305–325.

Samset, O. (1998). Stepwise Regression under Factor Sparsity using a new Stopping Criterion, *Technical Report 12*, Department of Mathematical Sciences. The Norwegian University of Science and Technology, Norway.

Samset, O. and Tyssedal, J. (1998a). Repeat and Mirror-Image Patterns and Correlation Structures of Plackett-Burman Designs, their Foldovers and Half-Fractions, *Technical Report 13*, Department of Mathematical Sciences. The Norwegian University of Science and Technology, Norway.

Samset, O. and Tyssedal, J. (1998b). Study of the Box-Meyer Method for finding Active Factors in Screening Experiments, *Technical Report 5*, Department of Mathematical Sciences. The Norwegian University of Science and Technology, Norway.

Tyssedal, J. and Samset, O. (1997). Analysis of the 12 Run Plackett-Burman Design, *Technical Report 8*, Department of Mathematical Sciences. The Norwegian University of Science and Technology, Norway.

Westfall, P. D., Young, S. S. and Lin, D. K. J. (1998). Forward Selection Error Control in the Analysis of Supersaturated Designs, *Statistica Sinica* **8**, 101–117.

Table 1: The 12 run Plackett-Burman design.

run	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}
1	+	-	+	-	-	-	+	+	+	-	+
2	+	+	-	+	-	-	-	+	+	+	-
3	-	+	+	-	+	-	-	-	+	+	+
4	+	-	+	+	-	+	-	-	-	+	+
5	+	+	-	+	+	-	+	-	-	-	+
6	+	+	+	-	+	+	-	+	-	-	-
7	-	+	+	+	-	+	+	-	+	-	-
8	-	-	+	+	+	-	+	+	-	+	-
9	-	-	-	+	+	+	-	+	+	-	+
10	+	-	-	-	+	+	+	-	+	+	-
11	-	+	-	-	-	+	+	+	-	+	+
12	-	-	-	-	-	-	-	-	-	-	-

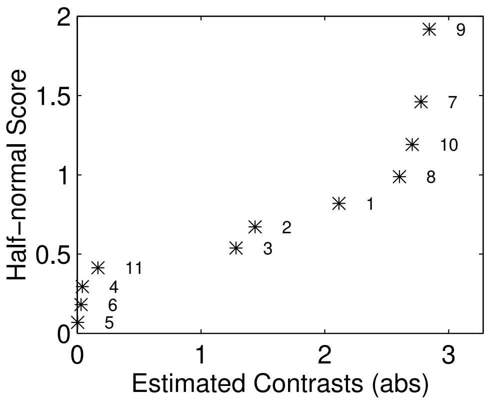

Figure 1: Half-normal plot of estimated main effect contrasts from the model $y = x_1 + 2x_1x_2 + 2x_1x_3 + \epsilon$ with $\epsilon \sim N(0, 0.25^2)$ using the 12 run PB design.

Table 2: The number of models to examine using best subsets regression.

No. of terms	PB(12)	PB(24)	PB(48)
1	66	276	1 128
2	2 145	37 950	635 628
3	45 760	3.47 e7	2.38 e8
4	720 720	2.37 e8	6.71 e10
5	8.94 e6	1.29 e10	1.51 e13
6	9.09 e7	5.81 e11	2.82 e15

Table 3: The number of models to examine using heredity least term regression.

No. of terms	PB(12)	PB(24)	PB(48)
1	11	23	47
2	165	759	3 243
3	1 705	17 963	163 231
4	15 510	384 307	7.51 e6
5	125 202	7.51 e6	3.19 e8
6	902 649	1.35 e8	1.26 e10

Table 4: The maximum number of models to examine using forward submodel regression.

No. of factors	No. of terms	PB(12)	PB(24)	PB(48)
1	1	11	23	47
2	3	165	759	3 243
3	6	990	10 626	97 290
4	10	3 300	88 550	1.78 e6
5	15	6 930	504 735	2.30 e7
6	21	9 702	2.12 e6	2.25 e8

Table 5: Identification of the correct model for example 1 for increasing σ. Response simulated 100 times for each value of σ.

			Model ranking (% cases)						
σ	Method	ΔR^2	1	2	3	4	5	6-10	> 10
0.5	Heredity		100						
	Forward	tuned	46	38	13	2	1		
	Forward	5.00	19	37	28	12	2		(2) 2
1.0	Heredity		93	4	2				1
	Forward	tuned	45	20	18	6	1	7	3
	Forward	5.00	13	21	22	15	4	10	(8) 15
1.5	Heredity		49	16	3	6	3	10	13
	Forward	tuned	21	10	8	12	4	7	38
	Forward	5.00	1	6	5	4	8	19	(27) 57

```
 Number of runs in design     :     12
 Number of factors in design  :     11
 Max number of terms          :      4
 Number of models examined    : 768692
 Number of rank warnings      :      0
 "Best model" criteria        : Residual MSe

Terms   R_sq  R_sqa     s  Terms in model
-----------------------------------------
    0   0.00   0.00  5.71
    1  57.16  52.88  3.92  3
    1  45.10  39.61  4.44  6*11
    1  41.18  35.30  4.59  8*10
    2  77.63  72.66  2.98  5*9 8*10
    2  76.99  71.88  3.03  3 3*4
    2  76.97  71.85  3.03  3 6*11
    3  93.68  91.31  1.68  5*9 8*10 10*11
    3  92.95  90.31  1.78  3 3*4 7*9
    3  92.09  89.13  1.88  3 7 6*11
    4  99.57  99.32  0.47  1 3 2*3 3*4
    4  97.01  95.30  1.24  8 5*9 8*10 10*11
    4  96.99  95.27  1.24  1*3 5*9 8*10 10*11
elapsed_time = 1711 seconds
```

Figure 2: Best subsets regression for example 1.

```
 Number of runs in design     :     12
 Number of factors in design  :     11
 Max number of terms included :      4
 Number of models examined    : 17392
 Number of rank warnings      :      0
 "Best model" criteria        : Residual MSe

Terms   R_sq  R_sqa     s  Terms in model
------------------------------------------
    0   0.00   0.00  5.71
    1  57.16  52.88  3.92  3
    1  15.17   6.69  5.51  1
    1   7.19   0.00  5.77  10
    2  76.99  71.88  3.03  3 3*4
    2  72.33  66.18  3.32  1 3
    2  65.24  57.52  3.72  3 3*2
    3  91.62  88.48  1.94  1 3 3*2
    3  85.07  79.47  2.59  3 3*2 3*4
    3  83.53  77.35  2.72  1 3 3*4
    4  99.57  99.32  0.47  1 3 3*2 3*4
    4  94.24  90.95  1.72  1 3 1*10 3*2
    4  94.06  90.66  1.74  1 3 1*6 3*2
elapsed_time = 40 seconds
```

Figure 3: Heredity least term regression for example 1.

```
 Number of runs in design     :    12
 Number of factors in design  :    11
 Max number of active factors :     4
 F-enter, F-remove            :  0.00  0.00
 Increase in R_sq             :  6.00  0.00
 Number of submodels examined :   562
 Number of models examined    :  2262
 "Best model" criteria        : Residual MSe

Active   R_sq  R_sqa     s  Terms in model
-----------------------------------------
     0   0.00   0.00  5.71
     1  57.16  52.88  3.92  3
     1  15.17   6.69  5.51  1
     1   0.00   0.00  5.71
     2  76.99  71.88  3.03  3 3*4
     2  72.33  66.18  3.32  1 3
     2  65.24  57.52  3.72  3 2*3
     3  91.62  88.48  1.94  1 3 2*3
     3  85.59  80.18  2.54  3 3*4 4*11
     3  85.07  79.47  2.59  3 2*3 3*4
     4  99.57  99.32  0.47  1 3 2*3 3*4
     4  99.57  99.06  0.55  1 8 1*4 1*8 2*4 2*8
     4  98.35  96.38  1.09  2 5 2*5 2*11 5*10 5*11
elapsed_time = 15 seconds
```

Figure 4: Forward submodel regression for example 1. $\Delta R^2_{\mathrm{ENTER}} = 6.0$.

Table 6: Half-fraction of the 24 run Plackett-Burman design.

run	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}	x_{18}	x_{19}	x_{20}	x_{21}	x_{22}
1	+	-	-	-	-	+	-	+	-	-	+	+	-	-	+	+	-	+	-	+	+	+
2	+	+	-	-	-	-	+	-	+	-	-	+	+	-	-	+	+	-	+	-	+	+
3	+	+	+	-	-	-	-	+	-	+	-	-	+	+	-	-	+	+	-	+	-	+
4	+	+	+	+	-	-	-	-	+	-	+	-	-	+	+	-	-	+	+	-	+	-
5	-	+	+	+	+	+	-	-	-	-	+	-	+	-	-	+	+	-	-	+	+	-
6	-	+	-	+	+	+	+	+	-	-	-	-	+	-	+	-	-	+	+	-	-	+
7	+	-	+	-	+	+	+	+	+	-	-	-	-	+	-	+	-	-	+	+	-	-
8	-	-	+	+	-	+	-	+	+	+	+	+	-	-	-	-	+	-	+	-	-	+
9	+	-	-	+	+	-	+	-	+	+	+	+	+	-	-	-	-	+	-	+	-	-
10	-	-	+	+	-	-	+	+	-	+	-	+	+	+	+	+	-	-	-	-	+	-
11	-	+	-	-	+	+	-	-	+	+	-	+	-	+	+	+	+	+	-	-	-	-
12	-	-	-	-	+	-	+	-	-	+	+	-	-	+	+	-	+	-	+	+	+	+

Table 7: Identification of the correct model for example 2 (columns 1, 2, 3 and 4) for increasing σ. Response simulated 100 times for each value of σ.

σ	Method	ΔR^2	Model ranking (% cases)						
			1	2	3	4	5	6-10	> 10
0.5	Heredity		100						
	Forward	tuned	85	12	3				
1.0	Heredity		56	12	11	7	2	7	5
	Forward	tuned	37	14	11	7	3	11	17
1.5	Heredity		17	10	11	3	5	11	43
	Forward	tuned	7	4	2	3	4	17	63

```
 Number of runs in design     :     12
 Number of factors in design  :     22
 Max number of terms included :      4
 Number of models examined    : 334610
 Number of rank warnings      :    132
 "Best model" criteria        : Residual MSe

Terms   R_sq  R_sqa     s  Terms in model
-----------------------------------------
    0   0.00   0.00  5.86
    1  49.56  44.52  4.37  3
    1  35.30  28.82  4.94  1
    1  22.73  15.00  5.40  15
    2  84.86  81.49  2.52  1 3
    2  71.55  65.23  3.46  6 6*18
    2  70.45  63.88  3.52  1 1*14
    3  95.44  93.74  1.47  1 3 1*10
    3  91.90  88.86  1.96  1 3 3*2
    3  90.96  87.57  2.07  1 3 1*19
    4  99.49  99.19  0.53  1 3 3*2 3*4
    4  97.84  96.60  1.08  1 3 1*10 3*2
    4  97.42  95.95  1.18  1 19 1*14 19*13
elapsed_time = 973 seconds
```

Figure 5: Heredity least term regression for example 2.

```
 Number of runs in design     :    12
 Number of factors in design  :    22
 Max number of active factors :     4
 F-enter, F-remove            :  0.00  0.00
 Increase in R_sq             :  7.00  0.00
 Number of submodels examined :  9109
 Number of models examined    : 34303
 "Best model" criteria        : Residual MSe

Active   R_sq  R_sqa     s  Terms in model
------------------------------------------
     0   0.00   0.00  5.86
     1  49.56  44.52  4.37  3
     1  35.30  28.82  4.94  1
     1  22.73  15.00  5.40  15
     2  84.86  81.49  2.52  1 3
     2  70.45  63.88  3.52  1 1*14
     2  69.66  62.91  3.57  3 2*3
     3  95.44  93.74  1.47  1 3 1*10
     3  93.86  91.56  1.70  13 13*19 19*22
     3  94.16  89.30  1.92  5 12 15 5*12 5*15
     4  99.49  99.19  0.53  1 3 2*3 3*4
     4  98.77  97.74  0.88  1 13 14 1*14 13*22
     4  99.08  96.64  1.07  17 18 19 8*17 8*18 8*19 17*19 18*19
elapsed_time = 275 seconds
```

Figure 6: Forward submodel regression for example 2. $\Delta R^2_{\text{ENTER}} = 7.0$.

Table 8: Identification of the correct model for example 2 (columns 1, 2, 5 and 6) for increasing σ. Response simulated 100 times for each value of σ.

			Model ranking (% cases)						
σ	Method	ΔR^2	1	2	3	4	5	6-10	> 10
0.5	Heredity		97	2	1				
	Forward	tuned	89	8	3				
1.0	Heredity		34	13	7	6	4	13	23
	Forward	tuned	28	12	4	10	8	16	22
1.5	Heredity		7	5	3	3	2	4	77
	Forward	tuned	5	2		3	3	9	78